

R E M A R K S

Claims 1-28, 31-34, 36-40, 43-51, and 57-65 were rejected under 35 USC 103 as being unpatentable over applicant's admitted prior art in view of Bothe, *Audio to Audio-Video Speech Conversion with the Help of Phonetic Knowledge Integration* (IEE, Jan 1997). Applicants respectfully traverse.

The Examiner summarizes what the Examiner believes to be applicants' admitted prior art, as art that

teaches a system (sic) with a TTS stream into a decoder, synthesizer, and compositor, along with a face model and FAPs into the compositor such that the output of the compositor is a synthesized audio/visual, wherein the timing of the audio and visual information is derived from the stream (sic).

Applicants respectfully disagree. In a nutshell, the Examiner's summary is phrased too expansively and consequently implies an admission that was not made. Specifically, applicants admitted that a face model is applied, *but not that the face model is part of the TTS stream*. Likewise, applicants admitted that FAPs are applied to a compositor (through the FRM), *but not that the FAPs are part of the TTS stream*. The Examiner's "along" connective word, however, can be viewed as including a teaching of a TTS stream with embedded FAPs; but as indicated above, that is **not** admitted by applicants to be in the prior art.

What the Bothe reference teaches is merely the creation of a string of phonemes from applied text and the conversion of the phonemes to video through a conversion that employs a codebook. Such a conversion is obviously not the same as the creation of video from FAP information that is applied, because it is the creation of video from phonemes that are developed.

In short, whereas the prior art creates a video from applied FAPs and a face model, Bothe creates a video by means for a totally different approach.

It is not clear what results from combining the two teachings (the admitted prior art and Bothe) because an artisan who is determined to do such combining would need to choose whether received FAPs are to be employed in order to create a video, or whether received FAPs should be discarded in favor of having the system develop phonemes from text (or speech, or a keyboard) and employ the phoneme to create a video in accord with the approach taught by Bothe. No skilled artisan would use both approaches because it

makes no sense to do so. Indeed, the only motivation for considering the Bothe reference is if, for some reason, the prior art method of employing received FAPs information is disfavored, and the effort is to replace the text-to-video approach of the admitted prior art with the approach that Bothe teaches. Applicants believe, however, that no skilled artisan would choose discarding received FAP information and, therefore, there is no cause for employing the teachings offered by Bothe.

More particularly relative to the first clause of amended claim 1, it defines  
a decoder responsive to an input signal stream comprising text  
commingled with FAP information, that separates the FAP information  
from the text, and develops phonemes from said text.

This clause thus excludes arrangements where FAP information is not commingled with text information in an input stream (singular). As for the Bothe reference, while it is true that it describes a decoder that develops phonemes from text, it is also true that the Bothe decoder (a) is not “responsive to an input signal stream comprising text commingled with FAP information,” and (b) does not separate “the FAP information from the text.”

The second clause of amended claim 1 specifies

a converter responsive to said decoder, that converts said phonemes to  
additional FAP information and outputs said additional FAP  
information combined with said FAP information separated by said  
decoder,

the Examiner’s summary does not even assert that applicants admitted to such a converter being in the prior art. As for a possible contribution by the Bothe reference, it is noted that it does not create FAP information from phonemes. Rather, it employs a codebook that maps a given, developed, phoneme to an optimal key-image sequence (see page 1635, right column, lines 7-8). It certainly does not add FAP information – or even the information that Bothe does develop – to the FAP information that is provided to the system. Therefore, the Bothe reference does not contribute anything material to the Examiner’s summary that would suggest the converter defined in amended claim 1.

Since neither the decoder nor the converter of amended claim 1 is taught or suggested by the admitted prior art (even as summarized by the Examiner) combined with Bothe, it is respectfully submitted that the rejection of amended claim 1 under 35 USC 103 is overcome.

Claim 2 is amended to include the limitation previously found in claim 5 (now deleted). With respect to claim 5, the Examiner stated that the asserted combination of the admitted prior art and Bothe

teaches basic FAPS (applicant's admitted prior art – non-viseme based, and other groupings of FAPS)”

Applicants respectfully disagree. The Bothe reference does not deal with FAPS at all, and applicants don't believe that they admitted that the prior art teaches the notion of a signal being generated – intended for a video synthesizer – which includes commands that are FAPS, but excludes viseme information. If the Examiner disagrees, applicants respectfully request that the Examiner cite the specific page and line where such an admission is allegedly made. In short, applicants believe that claim 2 defines a method that is not obvious in view of the admitted prior art in combination with the Bothe reference.

On a more general note, in connection with a significant number of the claims the Examiner rejected the claims with an explanation not unlike the explanation quoted above which, basically, asserts **without more**, that the cited combination of references teaches the claim. Respectfully, such an explanation is not helpful, because a rejection of a claim under 35 USC 103 inexorably arises from the fact that the Examiner believes that the combination of cited references teaches the claim. To restate it as the *explanation* of the rejection provides precious little by way of guiding the applicant as to the rationale, or the support, for the rejection.

Applicants endeavor herein to be as responsive as possible, based on whatever is gleaned from the Examiner's explanations. However, should applicants miss the mark and the Examiner remains unconvinced of the patentability of a claim because of specific teaching that is found in the cited art, applicants respectfully request that the Examiner cite specific page and line numbers (or col. and line numbers) in the references.

Claims 3-11 are deleted herein.

Claim 12 includes limitations that applicants believe are not suggested by combining the Bothe reference with the admitted prior art. Specifically, claim 12 specifies a decoder that is responsive to an input signal that comprises “signals representing audio **and embedded** video synthesis command signals” (emphasis supplied). From such an input signal the decoder creates two separate streams: one for

the audio signals, and the other for the video synthesis command signals. The Bothe reference clearly does not contemplate any such input signal, does not contemplate such a decoder, and does not create such two signal streams from a single input stream. The same is true for the admitted prior art. Accordingly, applicants respectfully submit that claim 12 is not obvious in view of the admitted prior art combined with Bothe.

Claims 13-30 depend on claim 12 and, hence, they are also not obvious in view of the admitted prior art combined with Bothe.

Additionally, at least some of the claims contain explicit additional limitations that are not suggested by the admitted prior art in combination with the Bothe reference. The following addresses a number of these claims.

Claim 16 specifies that *following* the separation of the input stream into two streams, the audio signal stream is converted to phonemes. That means that the video synthesis command signals stream is developed from something that precedes the development of phonemes. This, of course, is contrary to what the Bothe reference teaches, where information that dictates the video is derived from the phonemes.

Claim 20 specifies

a converter for generating additional video synthesis command signals, over and above said video synthesis command signals stream, from said phoneme signals and applying said additional video synthesis command signals generated by said converter to said video synthesizer, in addition to said video synthesis command signals stream being applied to said video synthesizer.

No such converter is in the admitted prior art, and no such converter is suggested in Bothe.

Claim 26 specifies that additional command signals are generated that are interpolated between command signals that are included within the input signal. Nothing like that is found in the admitted prior art, or in the Bothe reference. The Examiner appears to say that Bothe describes “interpolated signals including phoneme, timing, and command information ... data structure,” pointing to FIG. 12. Applicants respectfully submit that neither FIG. 12 of Bothe nor the text relating thereto teach or suggest that which the Examiner asserts. FIG. 12 merely shows that a key-vector may exist for each phoneme that is developed by the Bothe text-to-phoneme box. The text relative to FIG. 12 supports this analysis by stating

In the computer animation, the given phoneme input sequence  $\{Ph_i\}$  is mapped on a corresponding sequence of key-vectors  $\{Kp_i\}$  by an artificial neural network (NET-face [9]) as seen in fig. 12 ( $Ph_4$  has no key-image).”

Clearly, there is no notion of *interpolation* in the above-quoted text because interpolation means “to insert or introduce between other elements or parts” (Emphasis supplied) The American Heritage Dictionary, Houghton Mifflin Co., NY, 1982, and neither FIG. 12 nor the text associated with FIG. 12 suggests any interpolating or interposing.

Claim 27 more specifically defines the measure of the interposed command being an interpolation between the two adjacent pair of command signals. Since no command signals of the type defined in the claim are even present in the admitted prior art combined with Bothe, and since there are no interposed command signals of any kind in the admitted prior art combined with Bothe, it is not surprising that there is nothing to suggest the specific claim 27 definition of the measure of the interposed command signals.

With respect to claim 28 the Examiner also relies on FIG. 12 of Bothe, and it is tempting to assert that the key-vectors of FIG. 12 correspond to the command signals, that the phonemes of FIG. 12 somehow define a frame rate, that the video synthesizer generates images at a selected frame rate, and that an interpolation generates a command for each frame. It may be tempting, but it would be incorrect to do so. First, the phonemes do not come at any particular rate, and a close scrutiny of FIG. 12 bears this out (e.g., the distance between  $Ph_6$  and  $Ph_7$  is clearly greater than the distance between  $Ph_3$  and  $Ph_4$ ). Second, the key-vectors are not interpolated. Third, there no assurance that there is a key-vector that is associated with each phoneme (see, for example, phoneme  $Ph_4$ ). Hence, it is respectfully submitted that claim 28 is not obvious in view of the admitted prior art in combination with Bothe.

Claims 29 and 30 were indicated to be allowable, except for the fact that they depend on a rejected base claim.

Claim 31 is an independent method claim. Applicants respectfully submit that the remarks above apply to claim 31. Specifically, claim 31 specifies a step of

receiving an input signal that comprises signals representing audio and embedded video synthesis command signals;

The admitted prior art does not have such a signal, and clearly the Bothe reference does not have such a signal (and the Examiner has not explicitly pointed to such a signal).

Claim 31 also specifies

separating said input signal into an audio signal stream and a video synthesis command signals stream;

If the first step of the claim is not found, it is not surprising that this step is also not found, since it effectively requires the signal specified in the first step.

Amended claim 31 also specifies

synthesizing at least one image from said video synthesis command signals stream with aid of a FAP-based face model.

This step is not found in the admitted prior art, and it is certainly not found in the Bothe reference. Thus, applicants respectfully submit that claim 31 is not obvious in view of the admitted prior art combined with Bothe.

Claims 31-43 depend on claim 31.

It may be noted that claim 36 defines the further step of

of generating video synthesis command signals from said phonemes and said step of synthesizing is responsive to a combined command signals stream that includes said command signals developed in said step of separating and said command signals generated in said step of generating (emphasis supplied).

While it is true that the Bothe reference teaches generating video synthesis command signals from phonemes, the step of synthesizing is NOT responsive to a combined command signals stream that includes *both* the command signals developed in the step of separating and the command signals generated in the step of generating. Therefore, claim 36 includes an additional limitation that makes it even more patentable in view of the admitted prior art combined with the Bothe reference.

Claims 37-42 depend on claim 36.

Claim 38 defines a step of developing additional, interposed, command signals.

As discussed above, neither the admitted prior art nor the Bothe reference suggest such command signals.

Claim 39 defines the step of synthesizing at a "selected frame rate." Neither the admitted prior art nor the Bothe reference address the issue of a synthesis frame rate. The Examiner's explanation of the rejection points to FIG. 12 of Bothe, but applicants

respectfully submit that FIG. 12 does not describe, or even address, the question of a synthesizer's frame rate.

Claim 40 defines the measure of the interposed command lines in terms of interpolation between adjacent command signals. As indicated above, neither the admitted prior art nor the Bothe reference interpose command signals and, therefore, they don't even reach the issue of the measure of the interposed command signals.

Claims 41 and 42 were indicated to be allowable, except for the fact that they depend on a rejected base claim.

Amended claim 43 defines an apparatus. Applicants respectfully submit that the decoder/synthesizer defined in claim 43 is not described or suggested by the admitted prior art combined with the Bothe reference because it is responsive to an input stream (singular) that includes both text specification and explicit FAP information that is commingled with the text specification. Neither the admitted prior art nor the Bothe reference teaches such a signal.

Claims 44-56 depend on claim 43. Additionally, it is noted that at least a number of these claims contain explicit additional limitations that make the claims even more patentable over the admitted prior art in combination with the Bothe reference. The following addresses a number of those claims.

Claim 46 specifies that the FAP bookmarks in the input signal stream convey information about the identity of the FAP and the ultimate state of the FAP. In connection with this claim the Examiner merely asserts that the cited combination of admitted prior art and Bothe "teaches information about FAP's." Although FAPS are known in the art, it does not necessarily follow that the art suggests a signal that specifies text and embedded FAP bookmarks where the FAP bookmarks are characterized by a specification of the ultimate state of the FAP.

Similarly, applicants believe that the art does not teach a signal that specifies text and embedded FAP bookmarks where the FAP bookmarks are characterized by a specification of a duration measure for transiting to a specified state, as claim 48 defines, or the nature of the transition path, as claim 49 defines.

A similar argument applies to claims 50 and 51.

Claims 52-56 were indicated to be allowable, except for their dependence on a rejected base claim.

Claim 57 is independent. It defines a step of receiving an input that includes a text specification that is commingled with explicit FAP information and, as discussed above, neither the admitted prior art nor the Bothe reference employs such a signal.

One can assert that the Bothe reference develops two outputs (albeit not from the signal specified in the claim) where the first output is a synthesized voice, and phonemes are presented at a second output. However, that second output does not also develop FAP information. Hence, the Bothe reference is inapplicable. As for the admitted prior art, as indicated above applicants have not admitted that the prior art teaches the use of an input that includes a text specification commingled with explicit FAP information; and also have not admitted that the prior art teaches operating on such a signal to output "a synthesized voice at a first output, and phonemes as well as said FAP information at a second output." Therefore, applicants respectfully submit that claim 57 is not made obvious by the admitted prior art in combination with the Bothe reference.

Claims 58-65 are believed patentable for the reasons expressed above in connection with similar claims that depend on a different base claims.

Lastly, claims 66-70 were indicated to be allowable, except for their dependence of a rejected base claim.

In light of the above amendments and remarks, applicants respectfully submit that all of the outstanding claims overcome the rejection and, therefore, reconsideration and allowance are respectfully solicited.

Respectfully,  
Mark Beutnagel  
Ariel Fischer  
Joern Ostermann  
Yao Wang

Dated: 10/16/03

By Henry T. Brendzel  
Henry T. Brendzel  
Reg. No. 26,844  
Phone (973) 467-2025  
Fax (973) 467-6589  
email [brendzel@comcast.net](mailto:brendzel@comcast.net)